

# SuperNatural: a searchable database of available natural compounds

Mathias Dunkel\*, Melanie Fullbeck, Stefanie Neumann and Robert Preissner

Berlin Center of Genome Based Bioinformatics, 3D Datamining Group, Institute of Biochemistry, Charité–University Medicine Berlin, Monbijoustrasse 2, 10117 Berlin, Germany

Received August 15, 2005; Revised and Accepted October 24, 2005

## ABSTRACT

Although tremendous effort has been put into synthetic libraries, most drugs on the market are still natural compounds or derivatives thereof. There are encyclopaedias of natural compounds, but the availability of these compounds is often unclear and catalogues from numerous suppliers have to be checked. To overcome these problems we have compiled a database of ~50 000 natural compounds from different suppliers. To enable efficient identification of the desired compounds, we have implemented substructure searches with typical templates. Starting points for *in silico* screenings are about 2500 well-known and classified natural compounds from a compendium that we have added. Possible medical applications can be ascertained via automatic searches for similar drugs in a free conformational drug database containing WHO indications. Furthermore, we have computed about three million conformers, which are deployed to account for the flexibilities of the compounds when the 3D superposition algorithm that we have developed is used. The SuperNatural Database is publicly available at <http://bioinformatics.charite.de/supernatural>. Viewing requires the free Chime-plugin from MDL (Chime) or Java2 Runtime Environment (MView), which is also necessary for using Marvin application for chemical drawing.

## INTRODUCTION

The world is full of natural products, but only a few existing natural products are known and our understanding of the metabolome is fragmentary. Nature invented a universe of secondary metabolites as ‘defense compounds’ against enemies in predator–prey relationships. Concomitantly, strategies for

handling xenobiotics evolved, such as the multidrug resistance efflux pump and the cytochrome P450 monooxygenases (1,2). Tulp and Bohlin (3) hypothesize that when a natural compound occurs in unrelated species, it must have an important biological function, e.g. addressing a specific target, because fortuitous production of a particular compound by totally unrelated species is extremely improbable (3). About 200 000 natural compounds are currently known and many more will prove to be more than just ‘secondary metabolites’ (3). Even though combinatorial synthesis is now producing molecules that are drug-like in terms of size and property, these molecules, in contrast to natural products, have not evolved to interact with biomolecules (4). Natural compounds such as brefelidin A, camptothecin, forskolin and immunophilins often interfere with protein–protein interaction sites (5). Analysis of the properties of synthetic and natural compounds compared to drugs revealed the distinctiveness of natural compounds, especially concerning the diversity of scaffolds and the large number of chiral centers (6). This may be one reason why ~50% of the drugs introduced to the market during the last 20 years are derived directly or indirectly from natural compounds (7). Although most drugs on the market have a natural origin, their availability often remains unclear (8). The percentage of new non-synthetic chemical entities in the area of cancer remained at a yearly average of 62% over the period of 1981–2002 (9). Some marine natural products are either in or approaching Phase II/III clinical trials in cancer, analgesia, allergy and cognitive diseases (10). The chemical diversity of these compounds is tremendous and may offer inspiration for innovations in the fields of medicine, nutrition, agrochemical and life sciences (11).

## THE DATABASE

Several commercial databases and databases of rare compounds exist (12–14), but the SuperNatural Database is the first public resource containing 3D structures and conformers of 45 917 natural compounds, derivatives and analogues

\*To whom correspondence should be addressed. Tel: +49 30 450 528375; Fax: +49 30 450 528942; Email: mathias.dunkel@charite.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact [journals.permissions@oxfordjournals.org](mailto:journals.permissions@oxfordjournals.org)

purchasable from different suppliers. Currently, data from eight suppliers are available, but we plan to add further suppliers, compounds from which will be added on request (see 'List of Suppliers' on the SuperNatural Database website). The 2D structure of each compound, provided by the suppliers, was used to generate 3D structures (Discovery Studio, Accelrys Inc., <http://www.accelrys.com/dstudio>). Using a chemistry development kit (<http://almost.cubic.uni-koeln.de/cdk/>), fingerprints (966 bits, MACCS Keys) were calculated; each bit of a fingerprint represents functional groups (structural fingerprint). As a measure of 2D similarity we used the Tanimoto coefficient (15), which compares the bits of the structural fingerprints of two compounds. A Tanimoto coefficient of  $\geq 0.85$  indicates that a molecule has activities similar to a lead compound (16). For better coverage of the compounds and to ensure their flexibility during usage of the 3D-superposition algorithm, about three million conformers were evaluated (MedChem Explorer, Accelrys Inc., [http://www.accelrys.com/dstudio/ds\\_medchem](http://www.accelrys.com/dstudio/ds_medchem)). As a threshold for conformer generation, 20 kcal/mol as a relative maximum energy was set. This spacious threshold allows the user to find the best 3D superposition of two compounds even if they contain several rotatable bonds. The pre-computed fingerprints are stored in a MySQL-database on a web server, which is accessible via browser (see FAQ on the website for the database schema).

Owing to the immense structural diversity of natural compounds compared to synthetic compounds, an increased spectrum of therapeutic activities can be covered. Natural compounds can be classified by different criteria (see the classification list at 'Search via known compounds' on the SuperNatural Database website):

- (i) Classification by structural characteristics: alkaloid, amino acid, fatty acid, etc.
- (ii) Classification by functional aspects: vitamin, hormone, enzyme, etc.

To find desired natural compounds, a number of search options were implemented:

- As a starting point for screenings we compiled a searchable compendium of about 2500 well-known natural compounds characterized by a CAS-number (Chemical Abstracts), which is useful to cross-referencing other databases. This compendium contains systematic names, classification codes, empiric formulae, mixtures and synonyms (Figure 1A).
- Similarity searches based on fingerprints and Tanimoto coefficients are implemented in the SuperNatural Database (Figure 1B).
- Another way to perform a similarity search is the Marvin Applet, which allows the user to build or import a molecular structure and compare it with compounds of the SuperNatural Database (Figure 1C).
- Furthermore, an algorithm developed in our group enables 3D-superpositions of two compounds to be made. The algorithm compares all conformers of two compounds to find the best structural alignment (17) (Figure 1E).
- To identify possible applications, the user can search for similar drugs in the free drug database (SuperDrug Database) containing medical indications assigned by WHO (18).

About 300 natural compounds from the SuperNatural Database are identical to active ingredients of drugs, and 8% (3600) of the natural compounds are similar to essential marketed drugs with Tanimoto coefficients  $> 0.85$ . For each natural compound, information on different structural and chemical properties (DS Viewer, Property Calculator, <http://www.accelrys.com/dstudio>) such as number of chiral centers, estimated logp, surface area, etc. are precalculated and given in a separate 'FULL INFO' window (Figure 1D). For molecular visualization of the compounds, the user needs the free Chime-Plugin from MDL (available for Windows, SGI, Mac) or the Java2 Runtime Environment. Atomic coordinates of single or superimposed compounds are available for saving in Mol-format.

### PRACTICAL APPROACHES USING THE SIMILARITY SCREENING FUNCTION OF THE SUPERNATURAL DATABASE

A detailed review of various approaches to similarity searching was given by Willet *et al.* (19). Screenings for new bioactive natural compounds on the basis of chemical similarity to a known ligand depend on the similar property principle of Johnson and Maggiora (20). As an example, we performed a similarity screening in the SuperNatural Database with natural compounds that are known drugs, from clinical trials or lead compounds for drug development (Tables 1 and 2 and Supplementary Data) (21). Our investigations showed that the database contains compounds that have already been investigated in clinical trials for different diseases (Table 1 and 2 and Supplementary Data) and a great number of compounds with calculated 2D similarities of  $\geq 0.85$  to the lead compounds. The SuperNatural Database contains 289 natural compounds, which are already known as drugs. Owing to the immense structural and chemical variety of natural compounds, the coverage of a great spectrum of diseases is possible, which is confirmed by the ATC classifications of the drugs (see ATC classification in the category statistics on the SuperNatural website). There are 73 different ATC classes (three letter abbreviations) covered by these 289 natural compounds. The results show that the SuperNatural Database is an excellent source for finding bioactive natural products.

### AVAILABILITY

The database is publicly available at <http://bioinformatics.charite.de/supernatural>. The data will be updated twice a year.

### CONCLUSIONS AND FUTURE DIRECTIONS

The chemical diversity and unique properties of natural compounds provide a promising starting-point for developing innovations for scientific, medical and nutritional applications. The SuperNatural Database is a free resource with embedded screening functions for bioactive natural compounds. The extension of the database allows the scientific community simple access to a growing number of available natural compounds.

Adresse <http://bioinformatics.charite.de/supernatural/>

# Super Natural Database

[Home](#)  
[Compound Search](#)  

- via Known Compounds
- via Templates
- via Supplier/ID
- Build your own Structure
- For Similar Drugs

[FAQ](#)  
[List of Suppliers](#)  
[Statistics](#)  
[Contact](#)  
[Links](#)

## Compound Search via Known Compounds

**Type of Search**

Name  e.g. Vitamin A

CAS-Registry Number  e.g. 68-26-8

Synonym  e.g. 3,7-Dimethyl-9-(2,6,6-trimethyl...

Systematic Name  e.g. Retinol

Molecular Formula  e.g. C20-H30-O

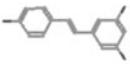
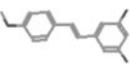
Molweight >  <  e.g. (> 280 < 300)

Chiral Centers >  <  e.g. (> 2 < 5)

LogP >  <  e.g. (> 3 < 5)

Classification

**Search**

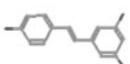
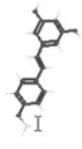
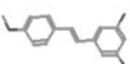
ID	Structure	Similarity	Supplier	3D Superposition
01502223		100.00	MicroSource	<input type="radio"/> Molecule 1 <input type="radio"/> Molecule 2 <input type="button" value="Go!"/>
01504044		87.41	MicroSource	<input type="radio"/> Molecule 1 <input type="radio"/> Molecule 2 <input type="button" value="Go!"/>

**PROPERTIES OF "01502223"**

NUMBER OF ROTATABLE BONDS:	2
NUMBER OF CHIRAL CENTERS:	0
NUMBER OF HYDROGEN BOND ACCEPTORS:	3
NUMBER OF HYDROGEN BOND DONORS:	3
LOG P:	3.014
SUM OF PARTIAL CHARGES:	0
SUM OF FORMAL CHARGES:	0
MOLECULAR DENSITY:	1.50
SURFACE AREA:	0.00

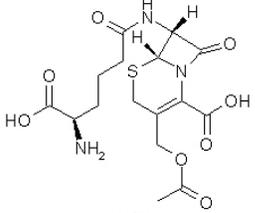
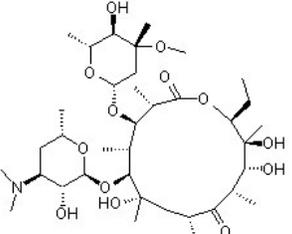
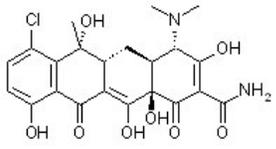
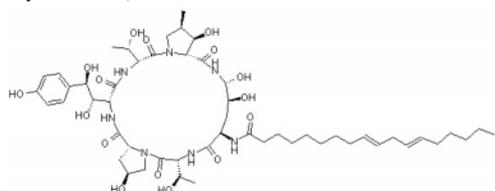
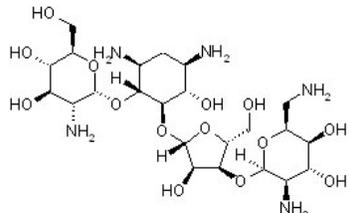
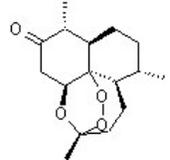
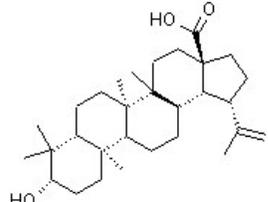
ID STRUCTURE 1: 01502223  
 ID STRUCTURE 2: 01504044

Number of atoms structure 1	: 17
Number of atoms structure 2	: 18
Superimposed atoms	: 17
Superimposed atoms with same type	: 17
Root Mean Square Distance	: 0.000

STRUCTURE 1	SUPERIMPOSED	STRUCTURE 2
		
ID: 01502223		ID: 01504044

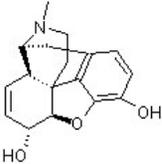
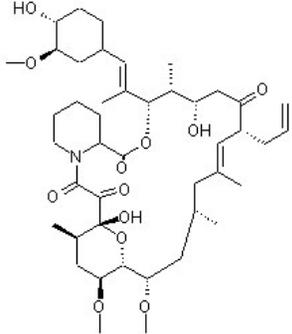
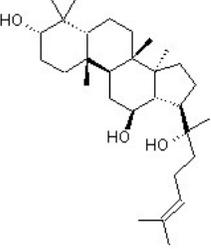
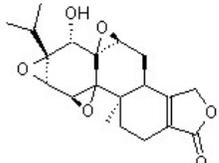
**Figure 1.** Screenshots of the web-interface of the SuperNatural Database. (A) Navigation frame and text query options for performing a search via known natural compounds. (B) Query results with the option for a 3D superposition. The 2D similarity query shows two compounds, which have a 2D similarity of 100.00 and 87.41 to the lead-structure. The compounds can be rotated (left mouse button), different display styles are available (right mouse button) and more detailed information concerning the properties of each structure can be obtained by use of the Properties button. Both compounds are available from the supplier MicroSource. (C) Screenshot of the Java applet Marvin, which allows upload or drawing of own structures for similarity searches in the SuperNatural Database. (D) Calculated properties for one structure. (E) Results of a 3D superposition. All conformations of both structures are superimposed and the best superposition is displayed. The table separately depicts the structures and the superposition of the corresponding conformations in the middle. The (superimposed) 3D structures can be saved by right clicking on the molecule. Also, information is given about the number of superimposed atoms and the root mean square distance.

**Table 1.** Well-known natural compounds (drugs, lead compounds for drugs or compounds in clinical trials) with antibacterial, antifungal, antiparasitic and antiviral effects and similar compounds (tanimoto  $\geq 0.85$ ) from the SuperNatural Database

Natural compound	2D structure	Similar compounds in SuperNatural (tanimoto $\geq 0.85$ )	Status (reference)
Antibacterial/J01* (antibacterials for sytemic use) Cephalosporin		20	Lead compound of cefalotin (21)
Erythromycin		15	Lead compound of flurithromycin (21)
(Oxy-, chlor-) tetracycline		27	Lead compound minocycline (21)
Antifungal/J02* (antimycotics for systemic use) Echinocandin B		4	Lead compound of caspofungin (22)
Antiparasitic/A07* (antidiarrheals, intestinal anti-inflammatory, anti-infective agents) P01* (antiprotozoals) Paromomycin		16	Active agent of paromomycin (23)
Artemisinin		7	Active agent of Artemisinin (24)
Antiviral/J05* (antivirals for systemic use) Betulinic acid		41	Phase I clinical trials (21)

\*Anatomical Therapeutic Chemical (ATC) classification code generated by the World Health Organization (WHO) describes the therapeutic subgroup (25).

**Table 2.** Well-known natural compounds (drugs, lead compounds for drugs or compounds in clinical trials) used in areas of neurological diseases, immunological or inflammatory processes and oncological diseases and similar compounds (tanimoto  $\geq 0.85$ ) from the SuperNatural Database

Natural compound	2D-structure	Similar compounds in SuperNatural (tanimoto $\geq 0.85$ )	Status (reference)
Neurological disease area/V03* (all other therapeutic products) Morphine		9	Lead compound of nalorphine (21)
Immunological, inflammatory/L04* (immunosuppressive agents) Tacrolimus		2	Active agent of tacrolimus (FK -506) (21)
Oncological disease area/L01* (antineoplastic agents) Protopanaxadiol		16	Phase I clinical trials (21)
Triptolide		18	Phase I clinical trials (21)

\*Anatomical Therapeutic Chemical (ATC) classification code generated by the World Health Organization (WHO) describes the therapeutic subgroup (25).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

This work is supported by the BMBF (grant contract: 0312705B) funded Berlin Center for Genome Based Bioinformatics (BCB). We would like to thank the companies for their permissions to use their natural compound libraries. Funding to pay the Open Access publication charges for this article was provided by Universitätsmedizin, Charité.

*Conflict of interest statement.* None declared.

## REFERENCES

- Schuler, M.A. and Werck-Reichhart, D. (2003) Functional genomics of P450s. *Annu. Rev. Plant Biol.*, **54**, 629–667.
- Del Sorbo, G., Schoonbeek, H. and De Waard, M.A. (2000) Fungal transporters involved in efflux of natural toxic compounds and fungicides. *Fungal Genet. Biol.*, **30**, 1–15.
- Tulp, M. and Bohlin, L. (2005) Rediscovery of known natural compounds: nuisance or goldmine? *Bioorg. Med. Chem.*, **13**, 5274–5282.
- Piggott, A.M. and Karuso, P. (2004) Quality, not quantity: the role of natural products and chemical proteomics in modern drug discovery. *Comb. Chem. High Throughput Screen.*, **7**, 607–630.
- Pommier, Y. and Cherfils, J. (2005) Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. *Trends Pharmacol. Sci.*, **26**, 138–145.

6. Feher, M. and Schmidt, J.M. (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.*, **43**, 218–227.
7. Vuorelaa, P., Leinonen, M., Saikkuc, P., Tammela, P., Rauhad, J.P., Wennberge, T. and Vuorela, H. (2004) Natural products in the process of finding new drug candidates. *Curr. Med. Chem.*, **11**, 1375–1389.
8. Koehn, F.E. and Carter, G.T. (2005) The evolving role of natural products in drug discovery. *Nature Rev. Drug. Discov.*, **4**, 206–220.
9. Newman, D.J., Cragg, G.M. and Snader, K.M. (2003) Natural products as sources of new drugs over the period 1981–2002. *J. Nature Prod.*, **66**, 1022–1037.
10. Newman, D.J. and Cragg, G.M. (2004) Advanced preclinical and clinical trials of natural products and related compounds from marine sources. *Curr. Med. Chem.*, **11**, 1693–1713.
11. Wessjohann, L.A., Ruijter, E., Garcia-Rivera, D. and Brandt, W. (2005) What can a chemist learn from nature's macrocycles?—a brief, conceptual view. *Mol. Divers.*, **9**, 171–186.
12. Qiao, X., Hou, T., Zhang, W., Guo, S. and Xu, X. (2002) A 3D structure database of components from Chinese traditional medicinal herbs. *J. Chem. Inf. Comput. Sci.*, **42**, 481–489.
13. Lei, J. and Zhou, J. (2002) A marine natural product database. *J. Chem. Inf. Comput. Sci.*, **42**, 742–748.
14. Fang, X., Shao, L., Zhang, H. and Wang, S. (2005) CHMIS-C: a comprehensive herbal medicine information system for cancer. *J. Med. Chem.*, **48**, 1481–1488.
15. Delaney, J.S. (1996) Assessing the ability of chemical similarity measures to discriminate between active and inactive compounds. *Mol. Divers.*, **1**, 217–222.
16. Martin, Y.C., Kofron, J.L. and Traphagen, L.M. (2002) Do structurally similar molecules have similar biological activity? *J. Med. Chem.*, **45**, 4350–4358.
17. Thimm, M., Goede, A., Hougardy, S. and Preissner, R. (2004) Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *J. Chem. Inf. Comput. Sci.*, **44**, 1816–1822.
18. Goede, A., Dunkel, M., Mester, N., Frommel, C. and Preissner, R. (2005) SuperDrug: a conformational drug database. *Bioinformatics*, **21**, 1751–1753.
19. Willet, P., Barnard, J.M. and Downs, G.M. (1998) Chemical similarity searching. *J. Chem. Inf. Comput. Sci.*, **38**, 983–996.
20. In Johnson, M.A. and Maggiora, G.M. (eds) *Concepts and Applications of Molecular Similarity*. Wiley, NY.
21. Butler, M.S. (2005) Natural products to drugs: natural product derived compounds in clinical trials. *Nature Prod. Rep.*, **22**, 162–195.
22. Detry, A. and Bart-Delabesse, E. (2005) Caspofungin: mode of action and therapeutic applications. *Rev. Med. Interne.*, in press.
23. Gupta, Y.K., Gupta, M., Aneja, S. and Kohli, K. (2004) Current drug therapy of protozoal diarrhoea. *Indian J. Pediatr.*, **71**, 55–58.
24. Rathore, D., McCutchan, T.F., Sullivan, M. and Kumar, S. (2005) Antimalarial drugs: current status and new developments. *Expert Opin. Investig. Drugs*, **14**, 871–883.
25. WHO (2003), The selection and use of essential medicines. Report of the WHO Expert Committee, 2002 (including the 12th Model list of essential medicines). *World Health Organ. Tech. Rep. Ser.*, **914**, 1–126.